Bayesian Hierarchical Modelling of Young Stellar Clusters

Javier Olivares Romero

Université Grenoble Alpes & Universidad Nacional de Educación a Distancia

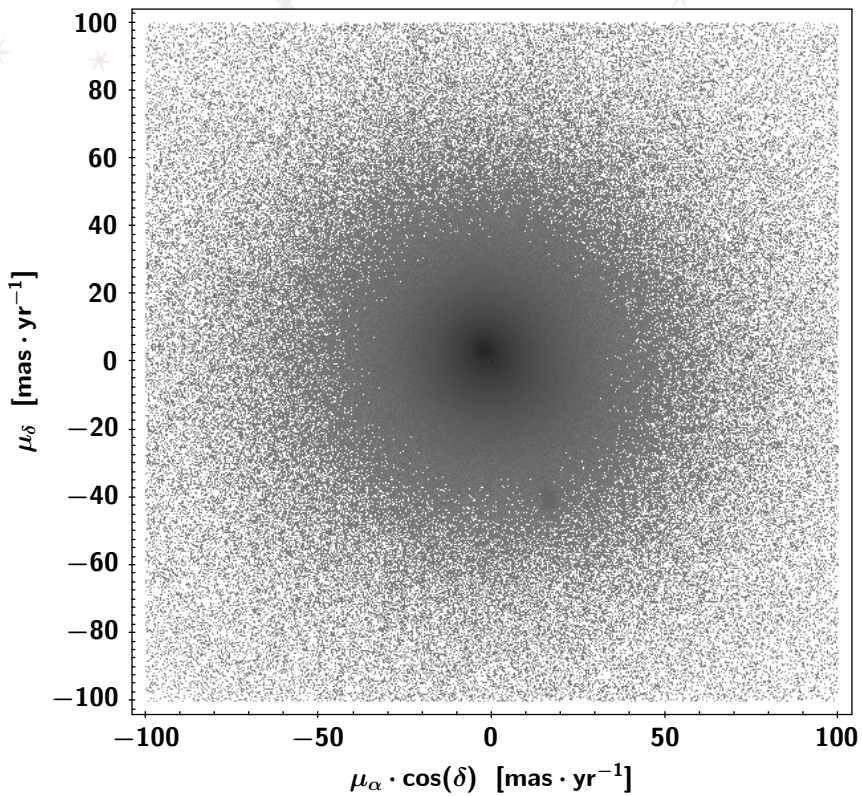
Supervisors:
Estelle Moraux & Luis Manuel Sarro Baro

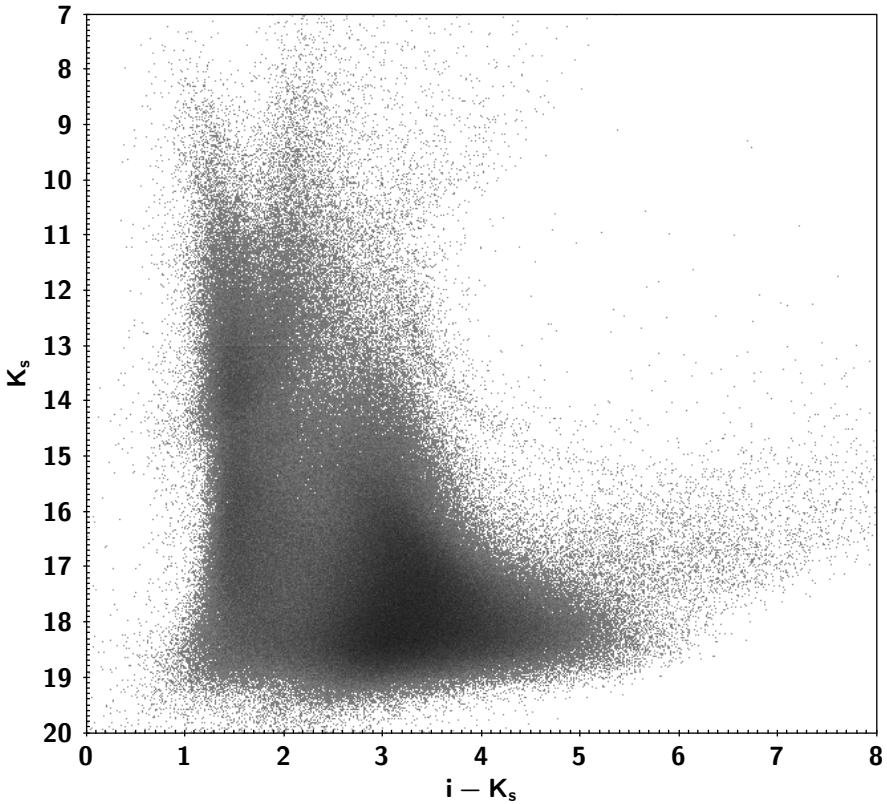
October 18, 2017

Stellar clusters: Origins

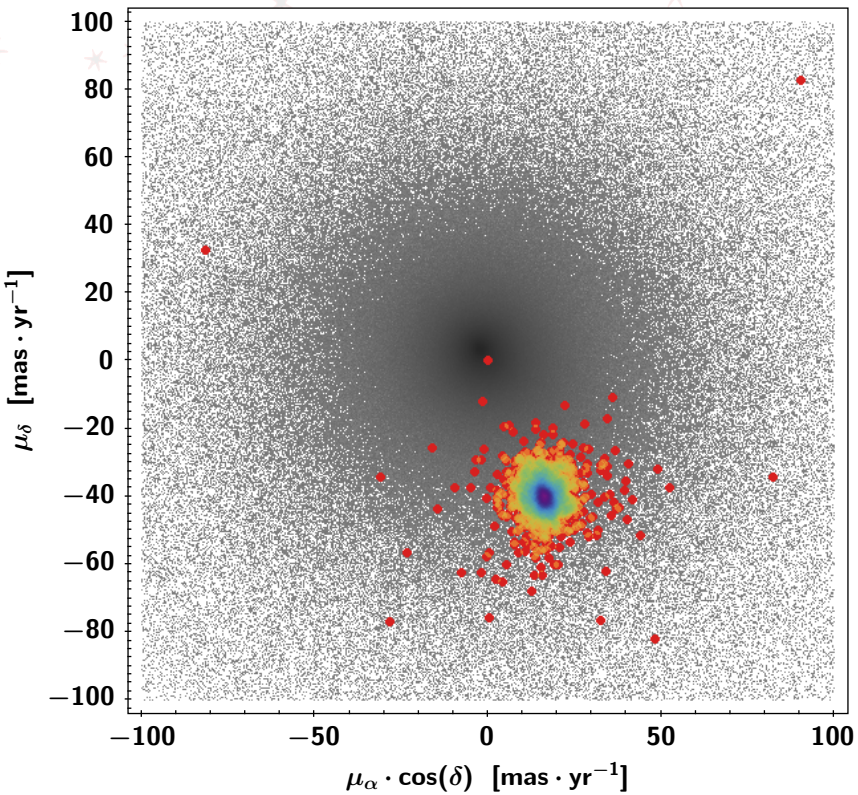
Stellar clusters: Origins

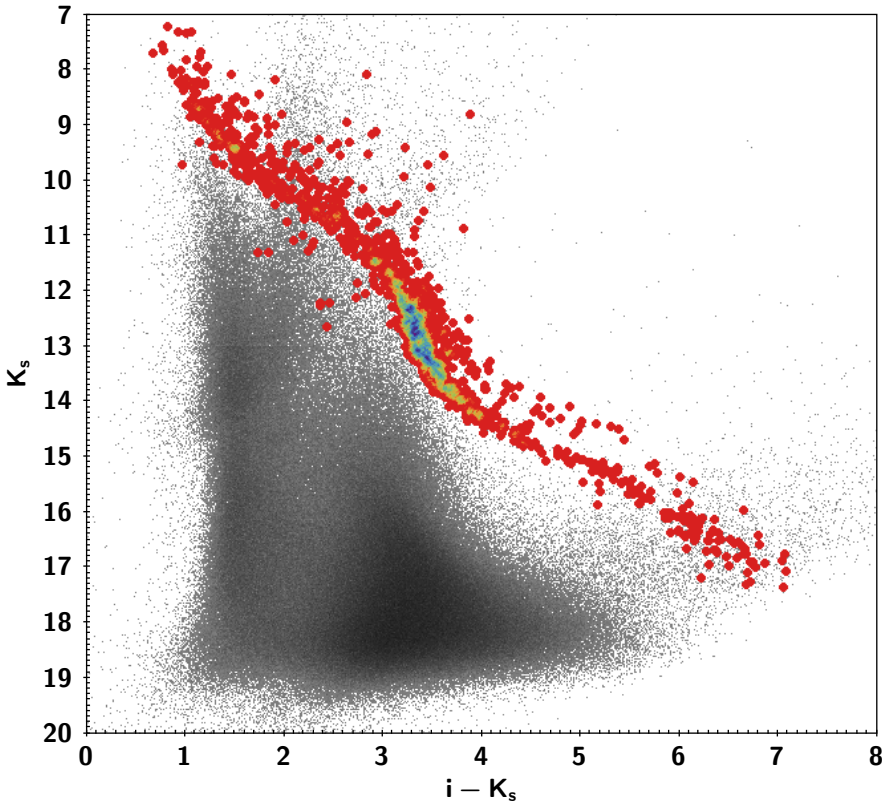
Stellar clusters: Observables





Stellar clusters: Observables





Objective

Obtain the statistical distributions of the cluster population

- Sky positions
- Proper Motions
- Apparent magnitudes

Simultaneously:

- Membership probabilities

Current approach

- BANYAN I (Malo et al. 2013)
- UPMASK (Krone-Martins & Moitinho 2014)
- Sarro (Sarro et al. 2014)
- BANYAN II (Gagné et al. 2015)
- Sampedro (Sampedro & Alfaro 2016)
- LACEwING (Riedel et al. 2017)

Caveats

- Classifiers only
- Homogeneous uncertainties
- Fully observed objects

Current approach

- BANYAN I (Malo et al. 2013)
- UPMASK (Krone-Martins & Moitinho 2014)
- Sarro (Sarro et al. 2014)
- BANYAN II (Gagné et al. 2015)
- Sampedro (Sampedro & Alfaro 2016)
- LACEwING (Riedel et al. 2017)

Caveats

- Classifiers only
- Homogeneous uncertainties
- Fully observed objects

Remedy

- Cluster distributions
- Heterogeneous uncertainties
- Missing value objects

Strategy

1. Create a new method.
2. Benchmark it.
3. Validate it
4. Evaluate it.
5. Discussion of the results.

- Bayesian Hierarchical Model
- Pleiades cluster
- Synthetic data
- DANCe DR2

Pleiades DANCe DR2 data set

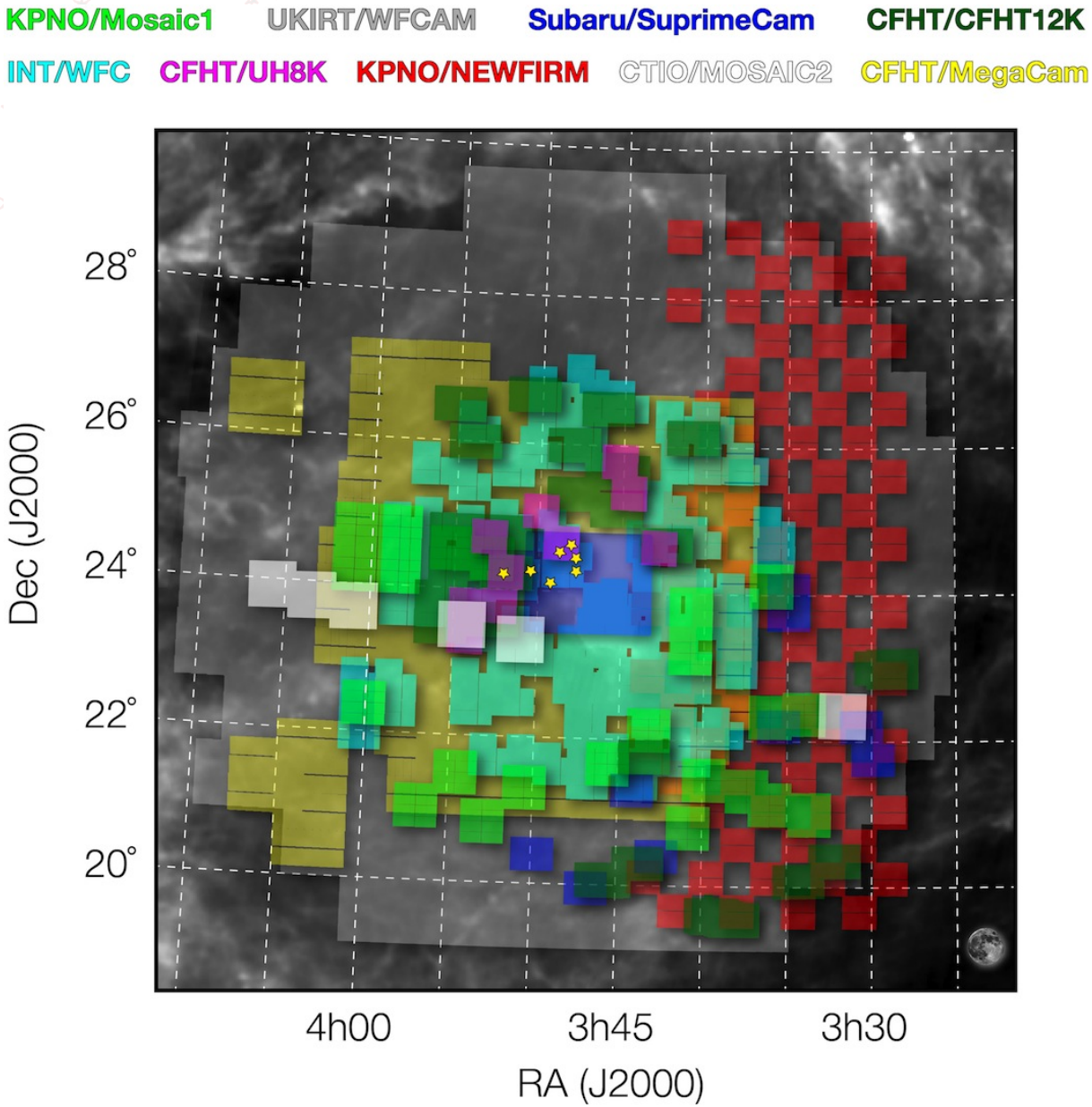
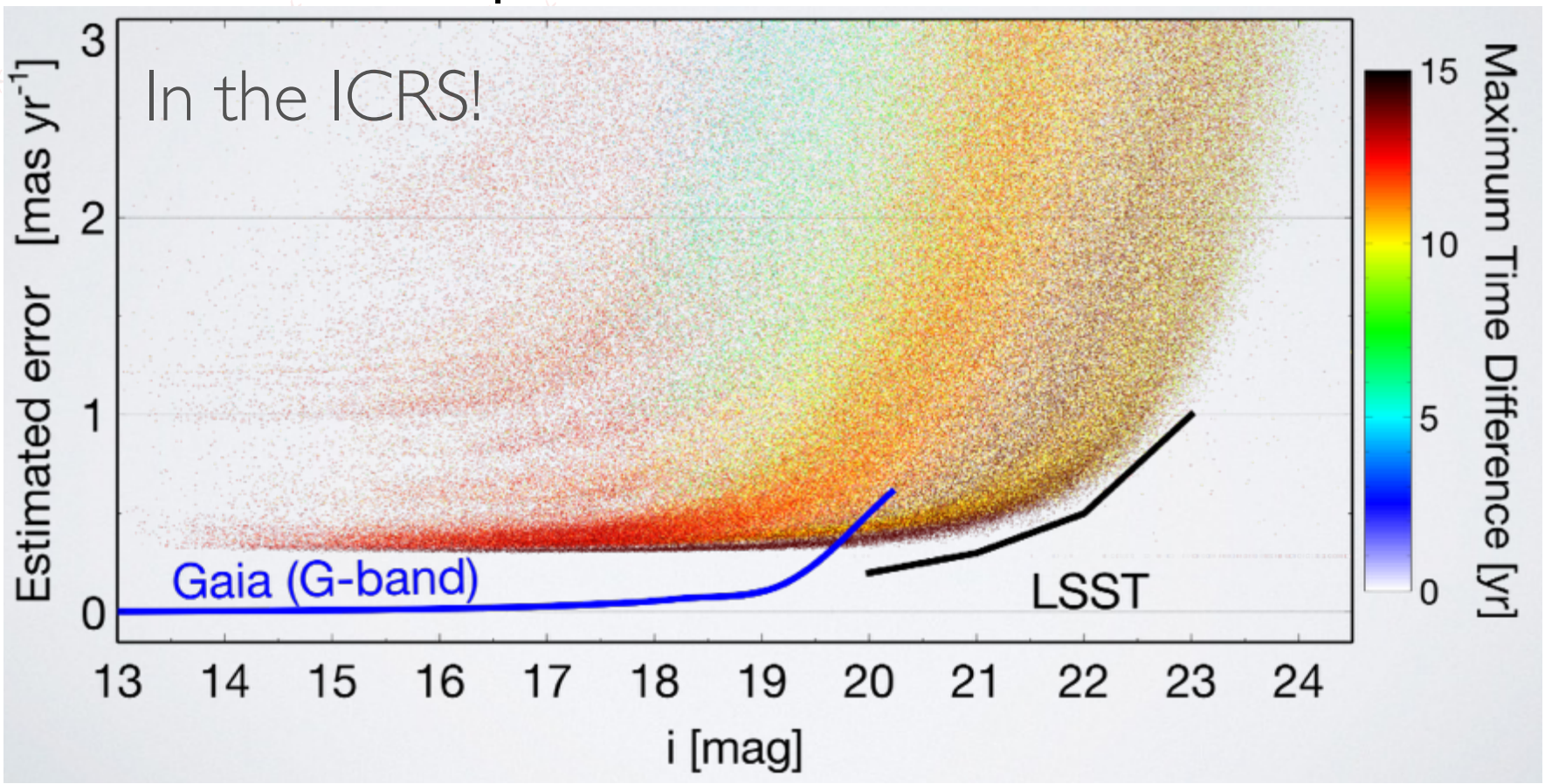


Image by Bouy et al. 2013

Pleiades DANCe DR2 data set

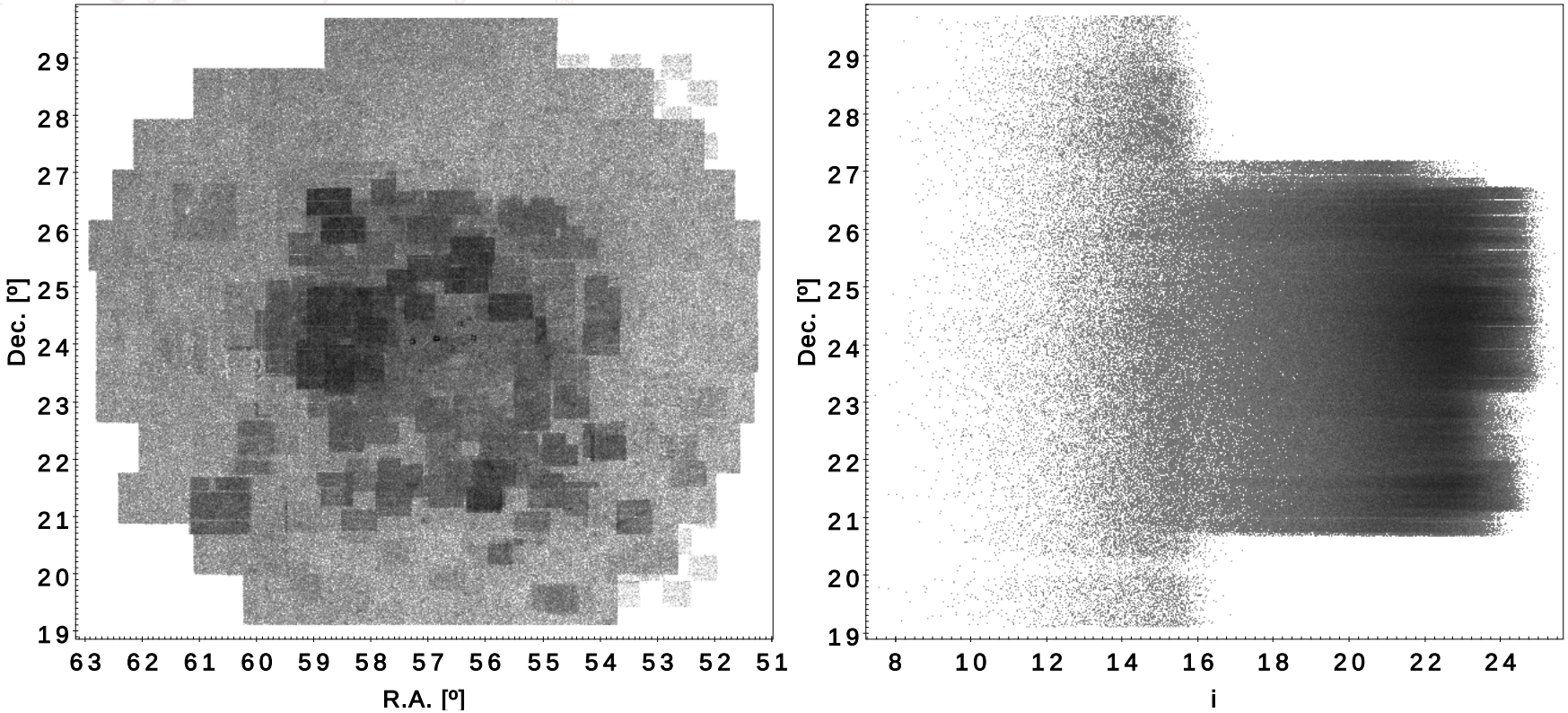
Complement Gaia at the faint end



Credit: Hervé Bouy

Pleiades DANCe DR2 data set

Complex probability of detection at all scales



Restricted data set

Reduce data set size

Computational constraints!

Selection

- Membership probabilities of Bouy et al. 2015
- Size: 10^5 objects.
- Probability threshold: 1×10^{-11}

Assumptions

- No cluster member outside the restricted sample

Overview of the method

Bayes' theorem

$$p(\boldsymbol{\theta}|\mathbf{D}, \mathcal{M}) = \frac{p(\mathbf{D}|\boldsymbol{\theta}, \mathcal{M}) \cdot p(\boldsymbol{\theta}|\mathcal{M})}{p(\mathbf{D}|\mathcal{M})}$$

Hierarchical model

Reduce prior subjectivity

Generative model

True values + Heteroscedastic data

Missing values

Types

- Deterministic
- Stochastic

Simplistic assumptions

- Ignorability
- Missing at random

Details of the method

Selection of observables

- Proper motions and $i - Ks, Y, J, H, K_s$
- Sky positions (independent analysis)

Generative model

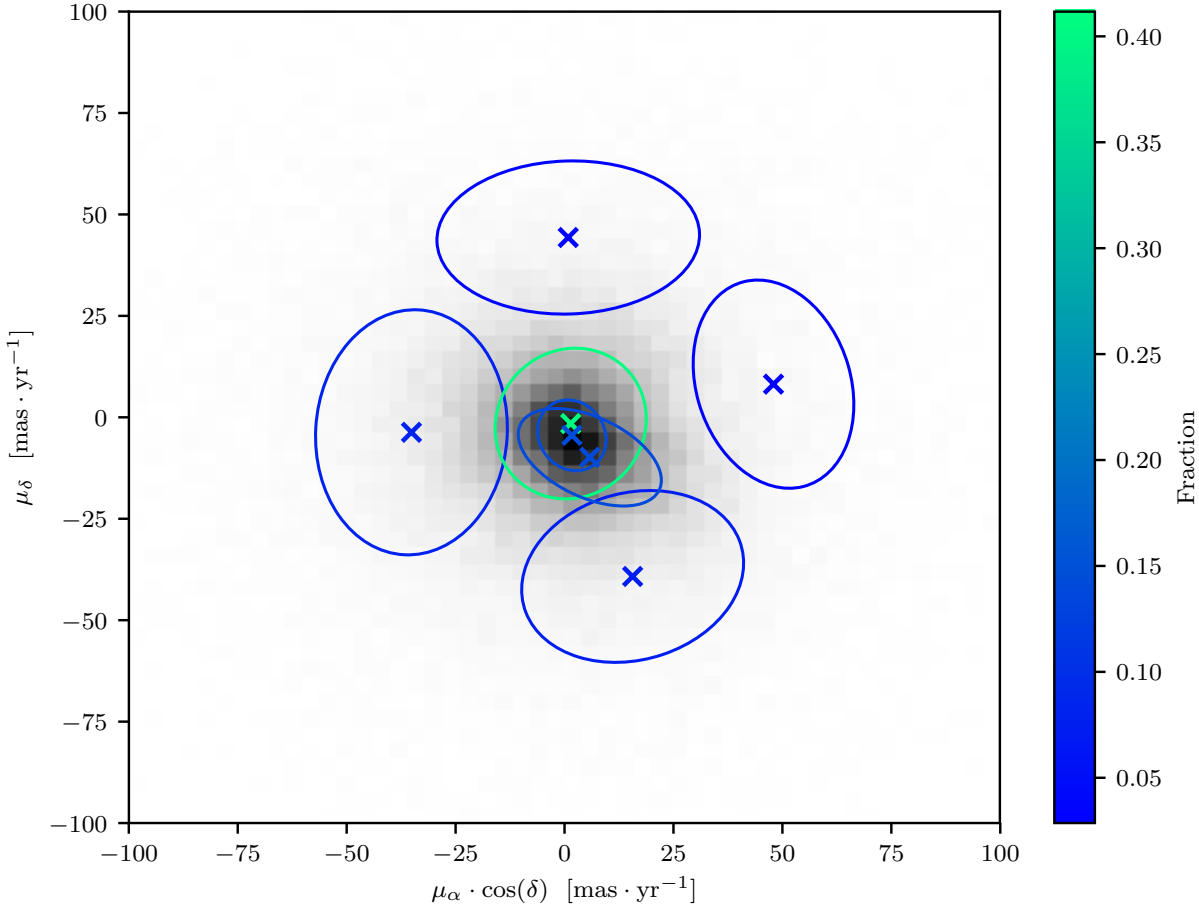
Field + Cluster

Assumptions

- Observables of Sarro et al. 2014 are correct.
- I hold the field fixed.
- Photometry independent of proper motions.
- Populations: Single stars and Equal-Mass Binaries (EMB).

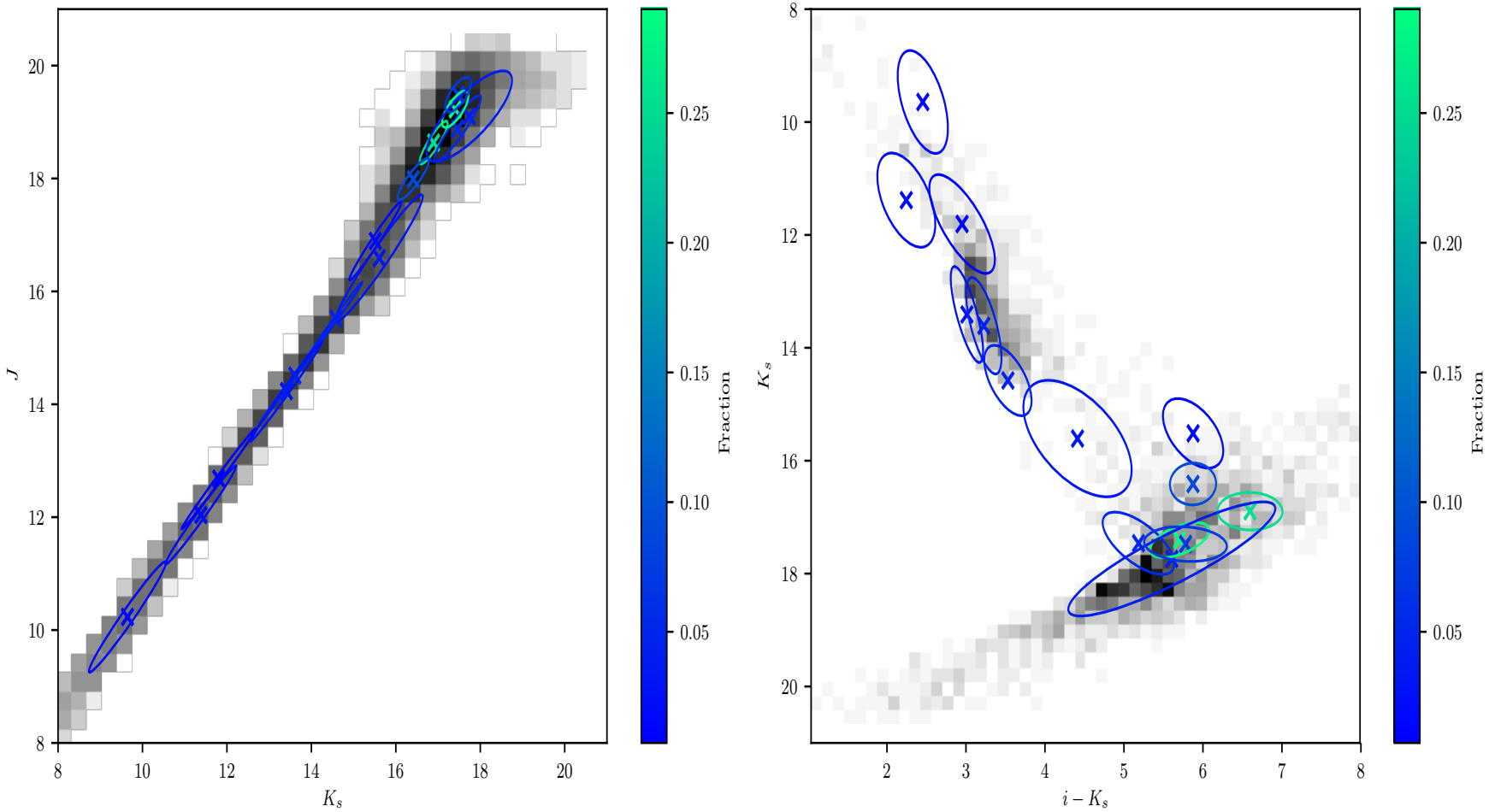
Field: Proper Motions

Gaussian Mixture Model + Uniform with 8 components (BIC)
Maximum-Likelihood Estimate



Field: Photometry

Missing values Gaussian Mixture Model with 14 components (BIC)
Maximum-Likelihood Estimate



Cluster model

Single stars + Equal-Mass binaries $\rightarrow$ 85 parameters

Proper motions

- GMM
  - Singles: 4 components
  - EMB: 2 components

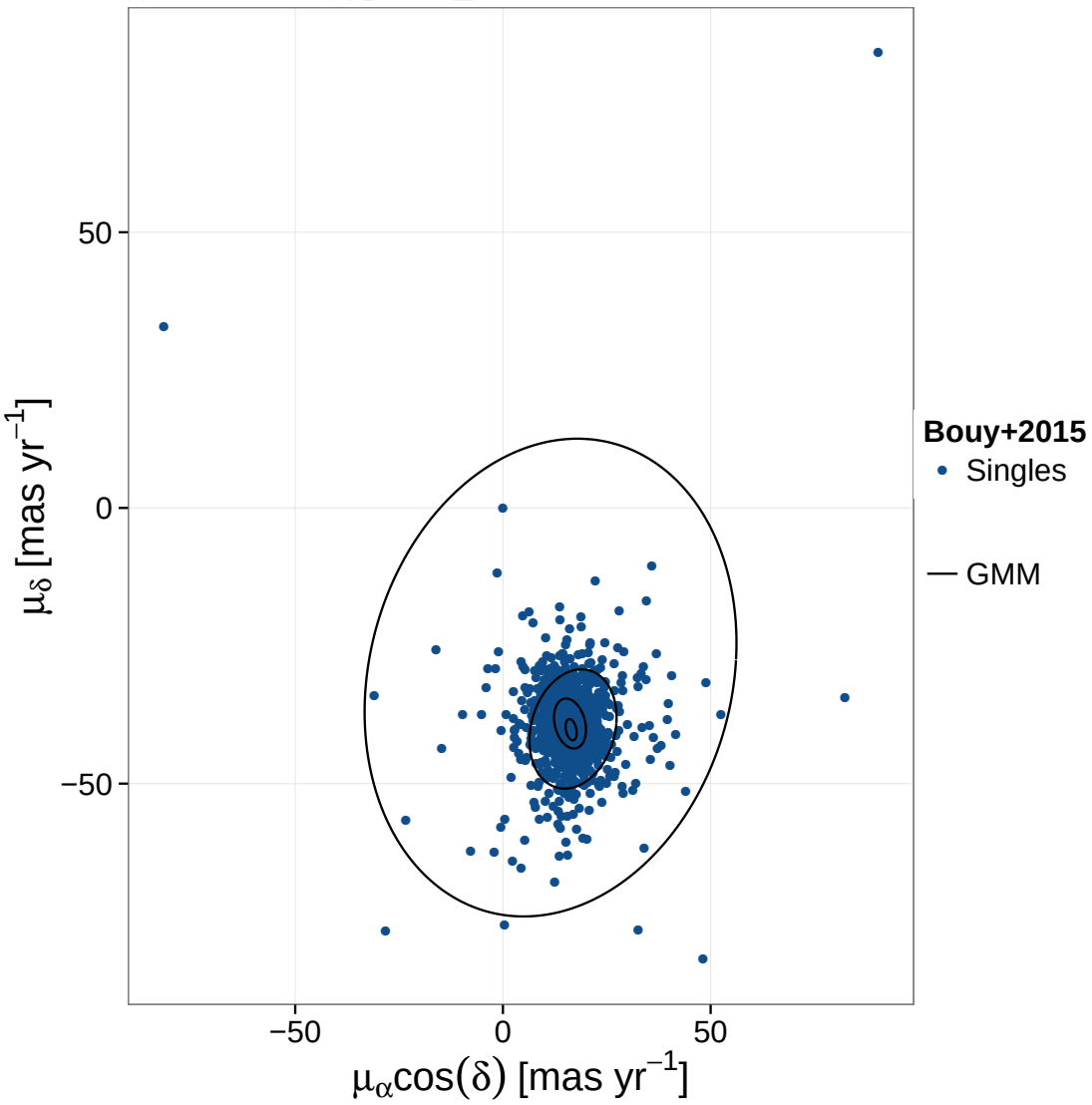
Photometry

- True $i - K_s$ distribution
  - GMM: 5 components
- Splines: Y, J, H, K_s
- Intrinsic dispersion
  - 5D Gaussian

Assumptions

- Number of GMM components given by BIC
- Splines parameter: $i - K_s$
- Splines knots remain fixed

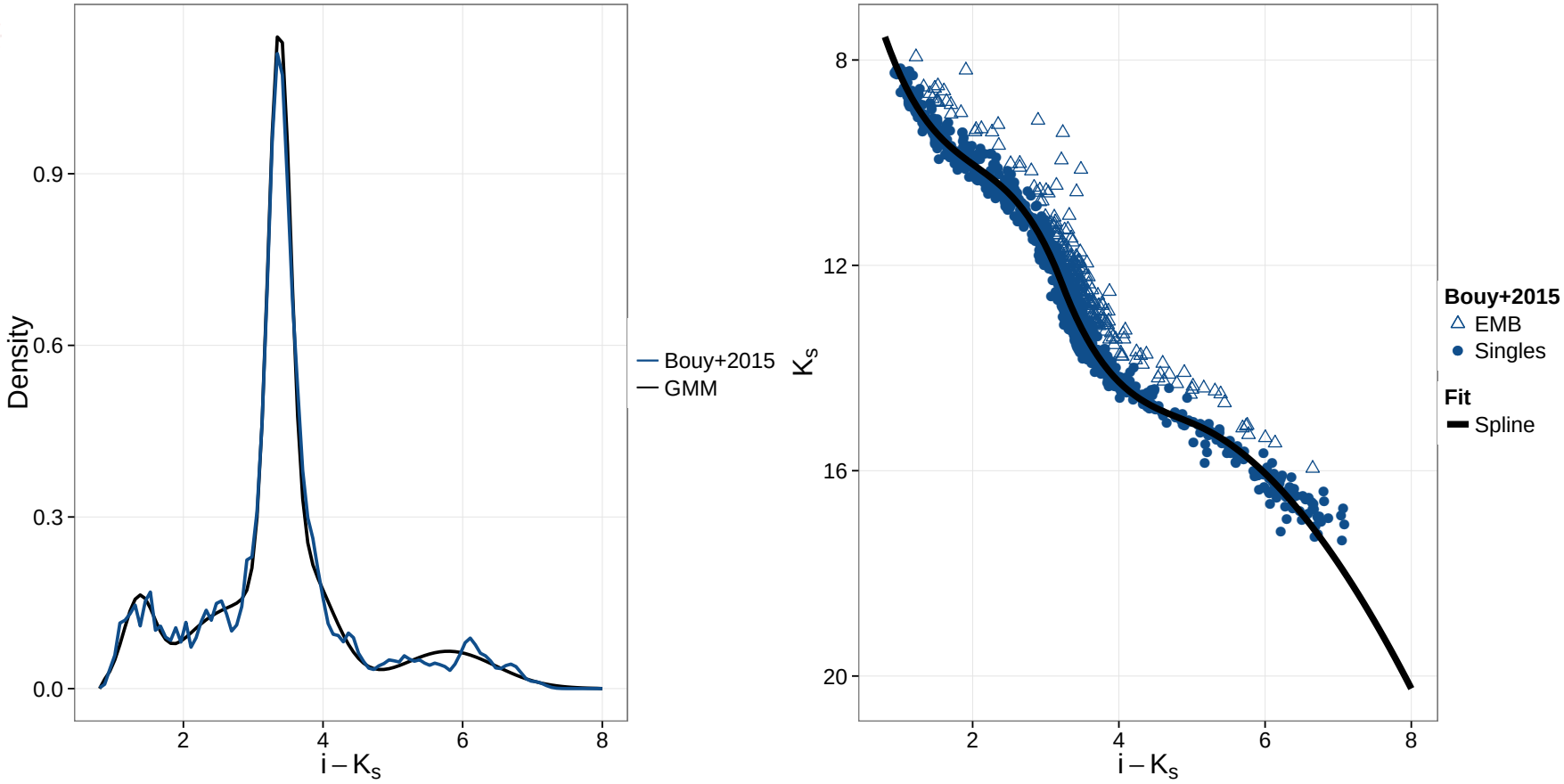
Cluster proper motions model



Gaussian Mixture

- Singles: 4 components
- EMB: 2 componentes

Cluster photometric model



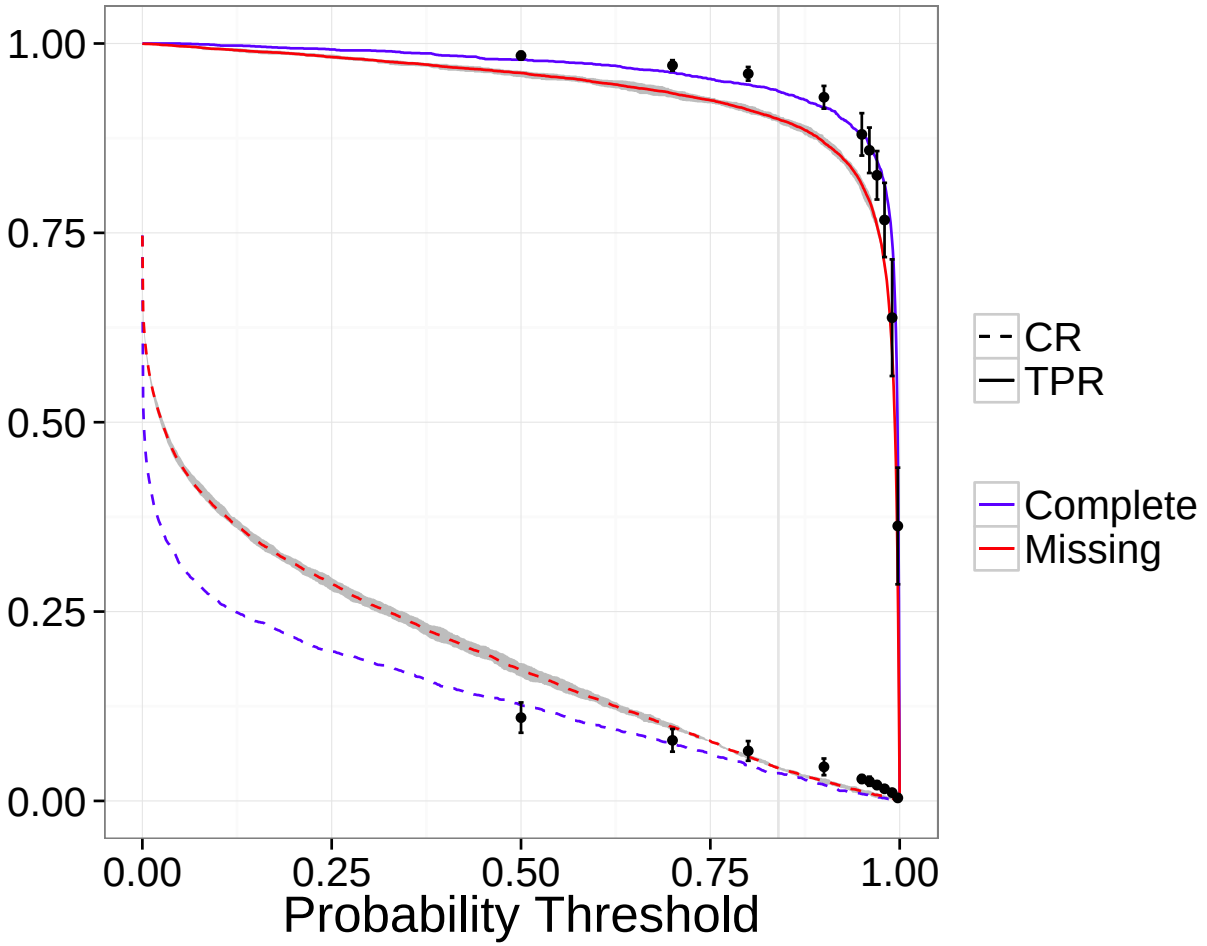
Obtaining the cluster distributions

High-Performance Computing

- Diverse infrastructures: UCA,UGA,CAB,UB
- Hybrid MPI-Multithreading approach
- Heuristic Particle Swarm Optimiser
- Markov Chain Monte Carlo
  - Convergence

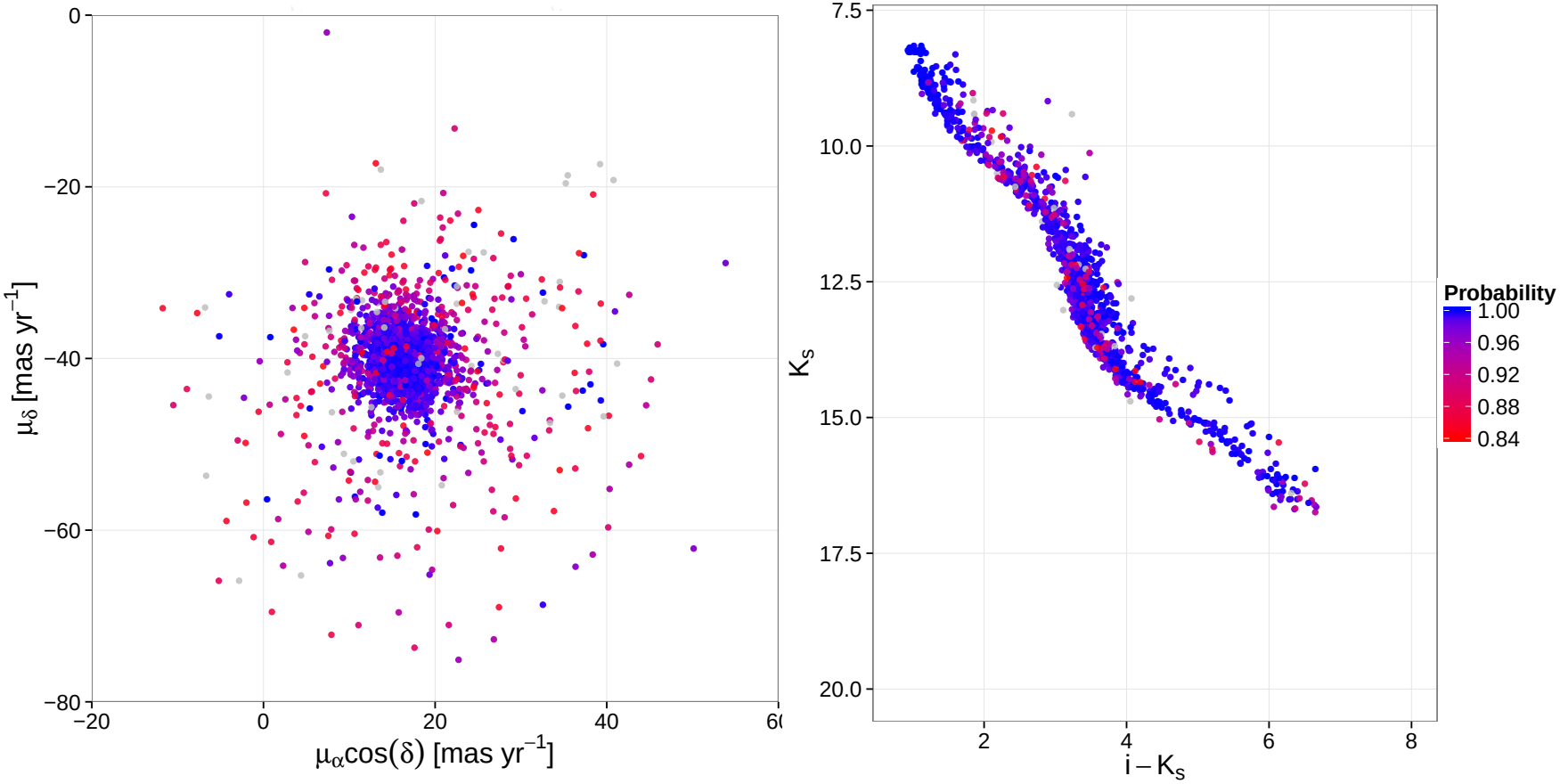
Validation on synthetic data

- Optimum Threshold = 0.84
- Contamination Rate = $4.3 \pm 0.2\%$
- True Positive Rate = $90.0 \pm 0.05\%$
- Expected value of CR = $5.8 \pm 0.2\%$



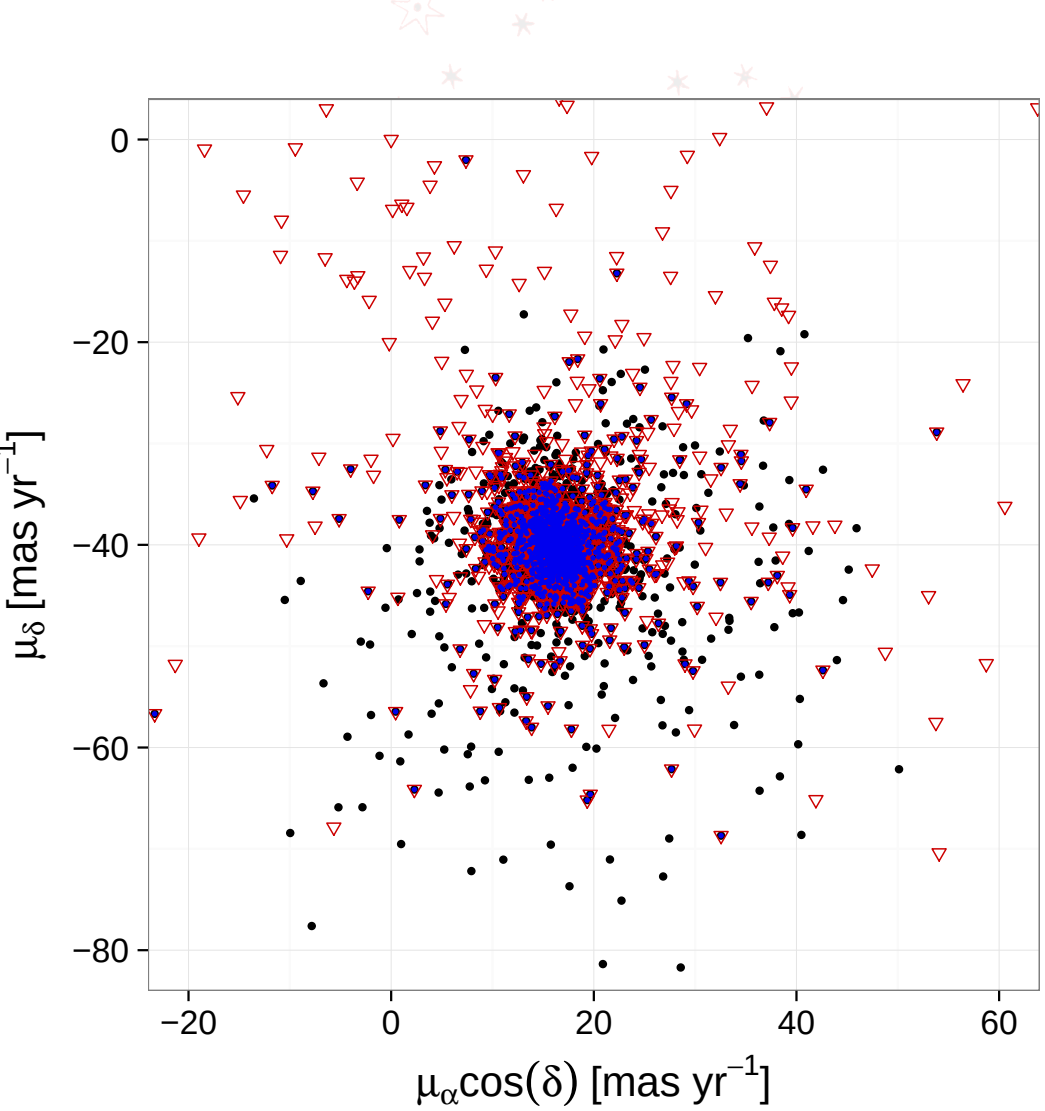
Results on the DANCe DR2

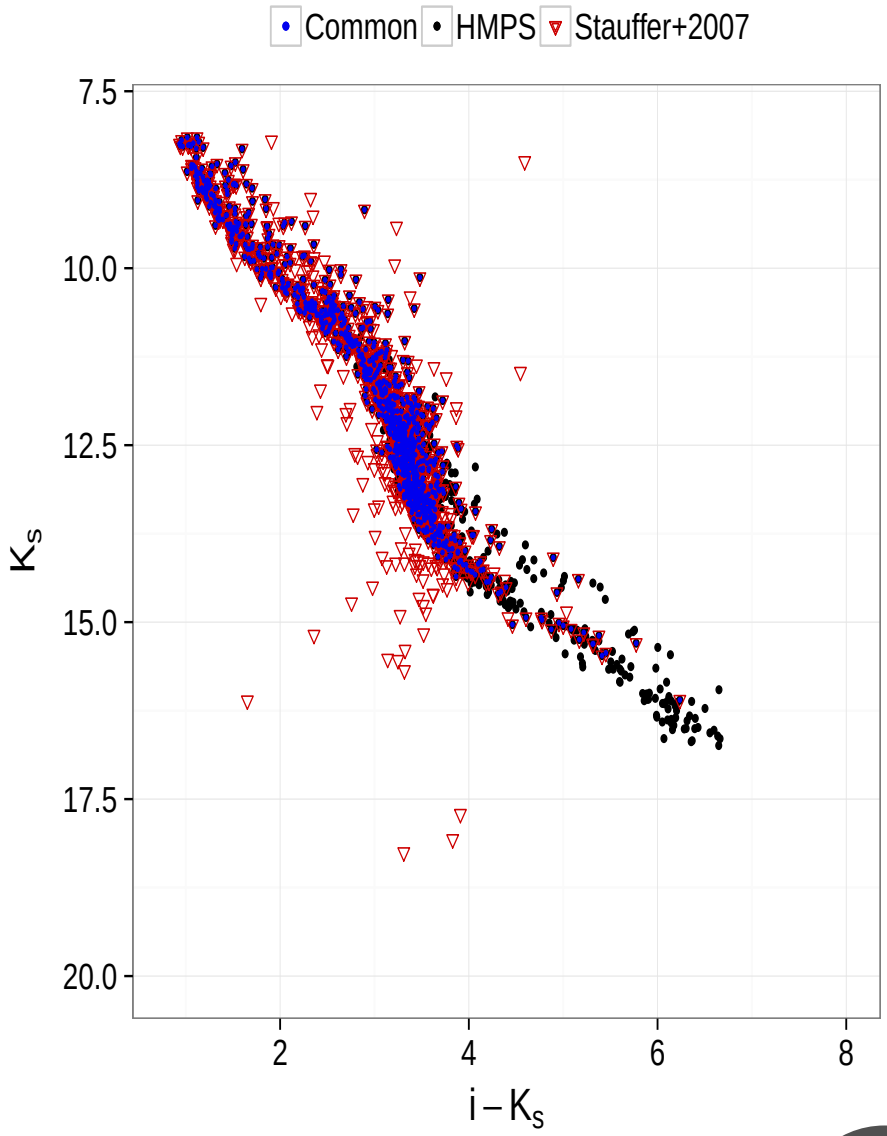
- Cluster membership probabilities $\rightarrow$ HMPS



Evaluation: Stauffer et al. 2007

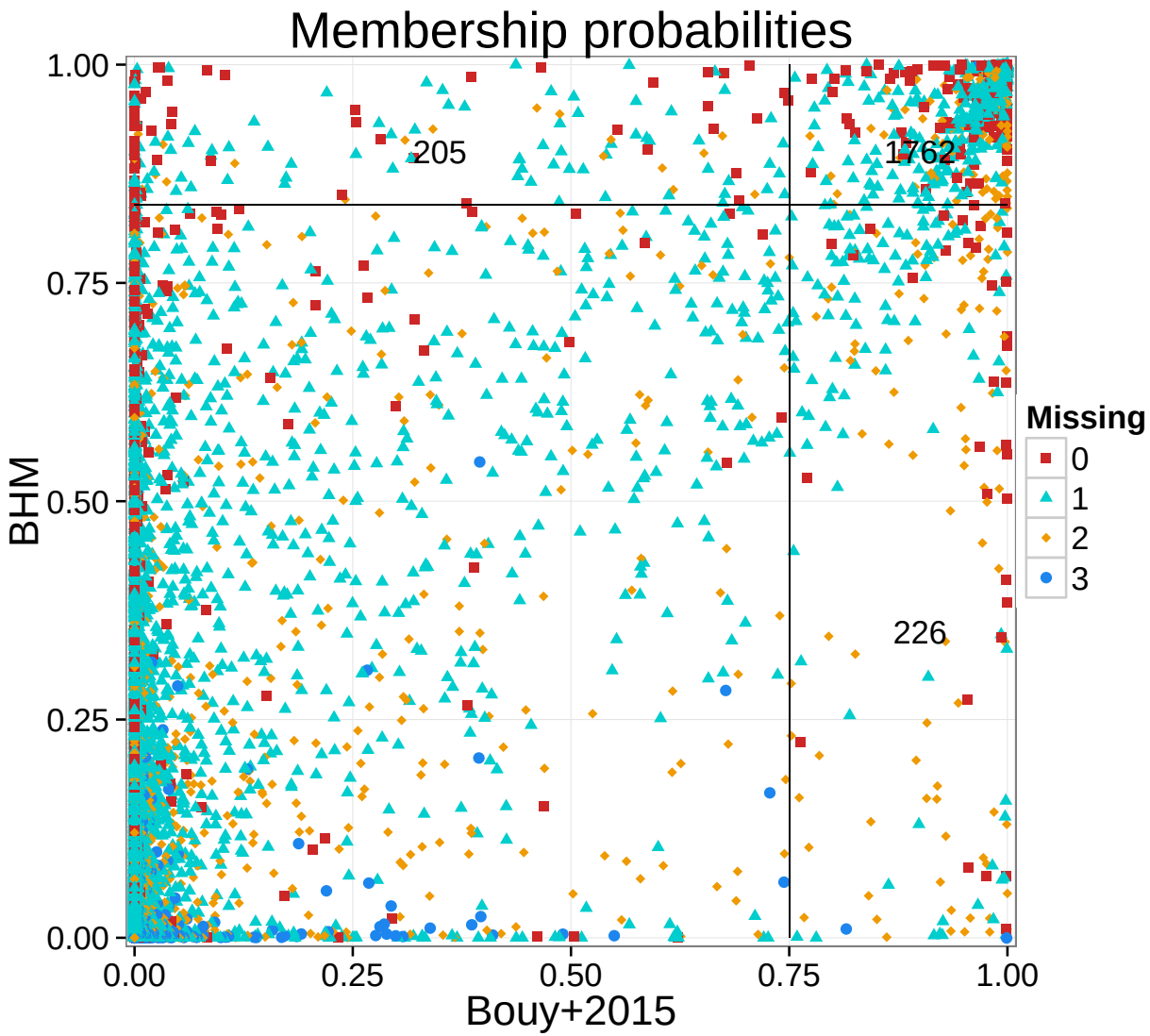
- Agreement: 80%



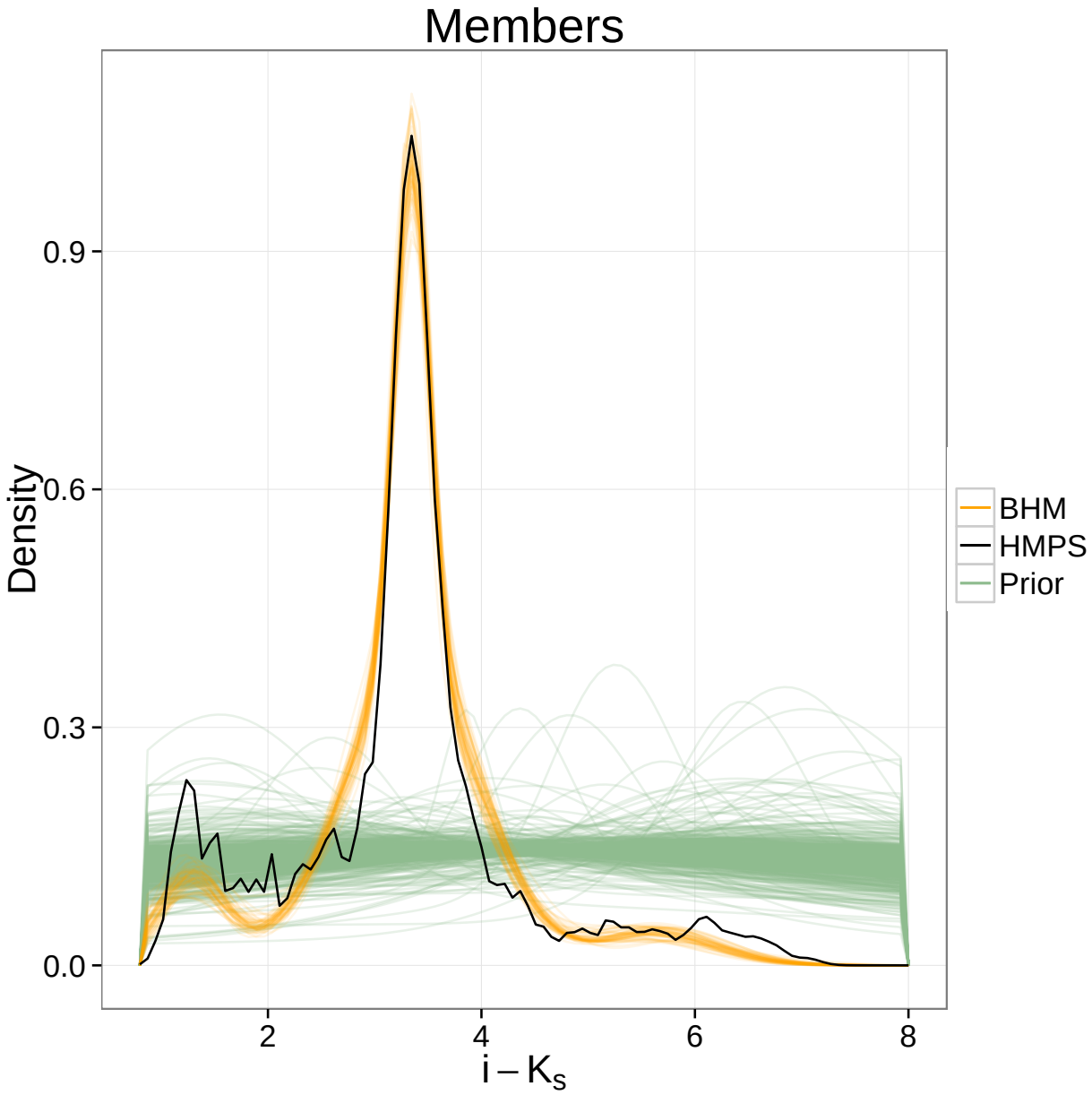


Evaluation: Bouy et al. 2015

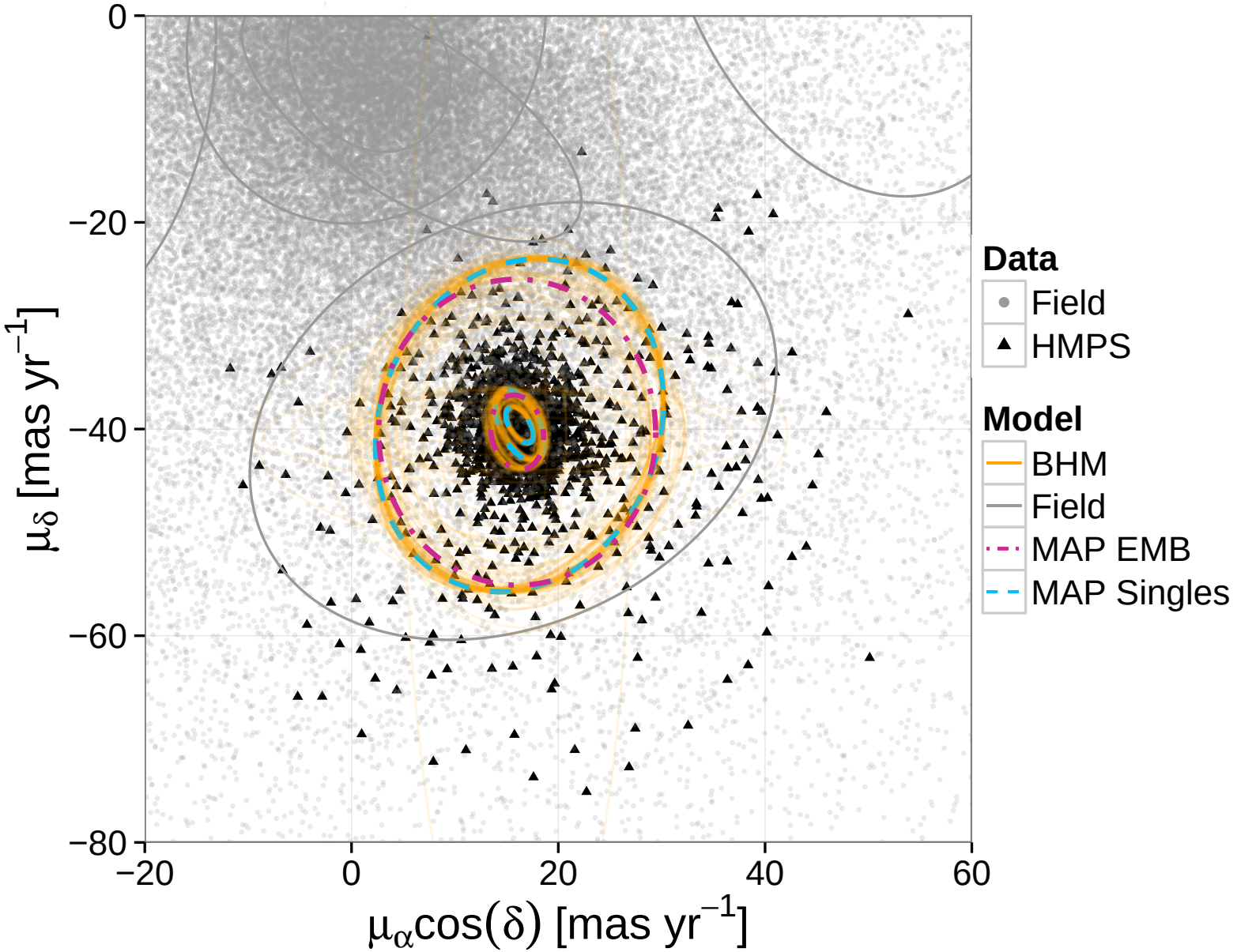
- Agreement: 88%



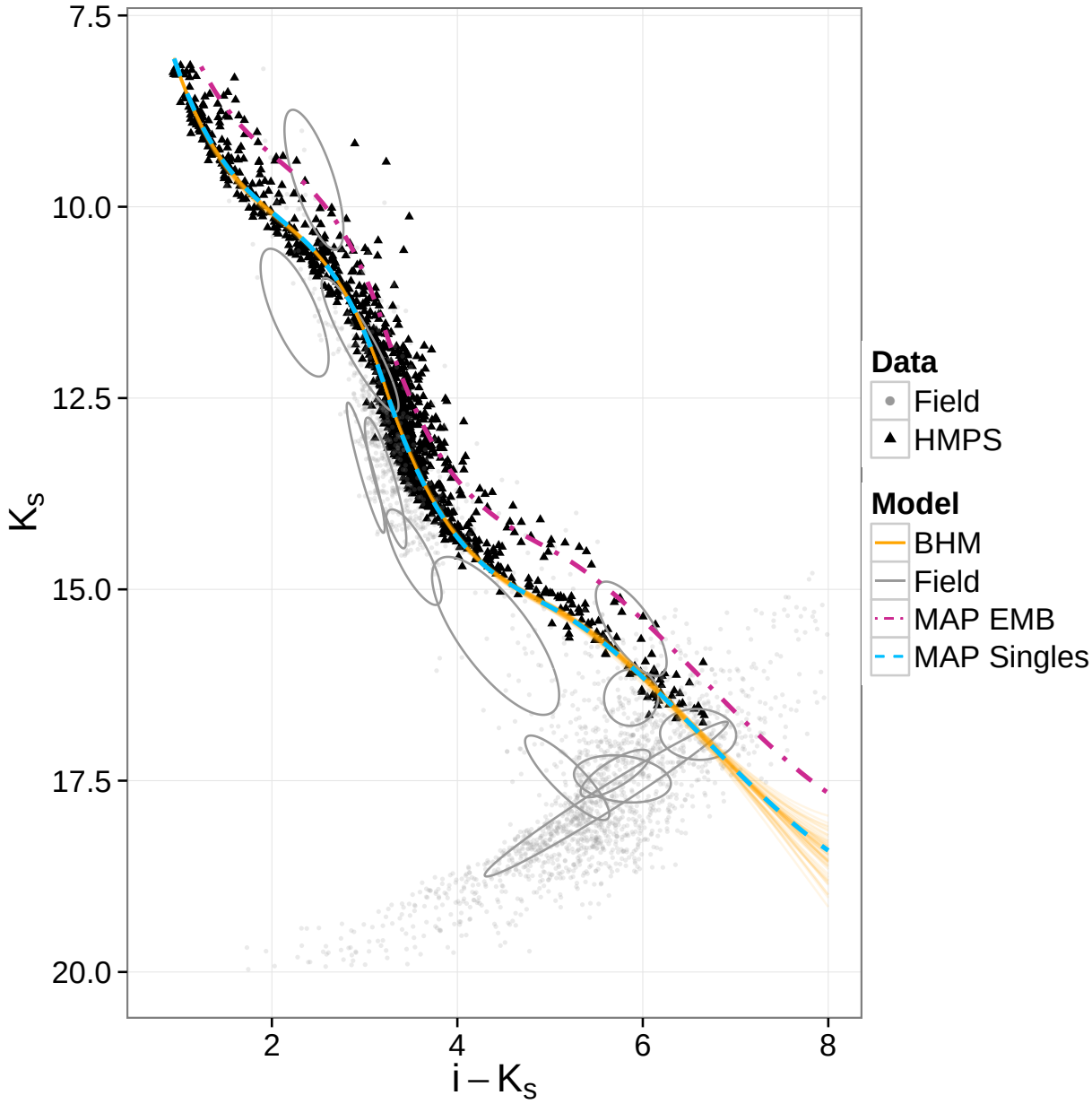
The statistical distributions: Colour index



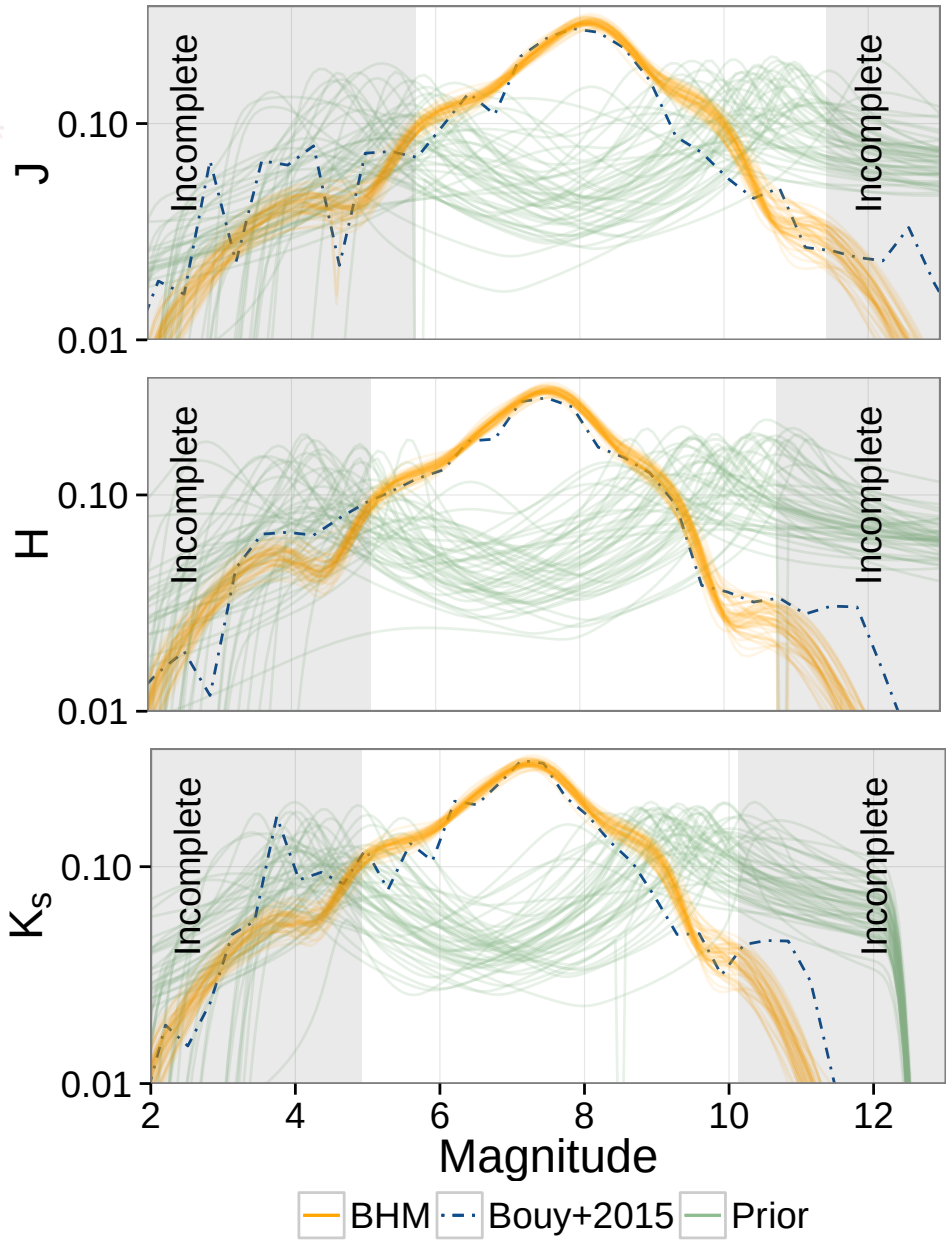
The statistical distributions: Proper motions



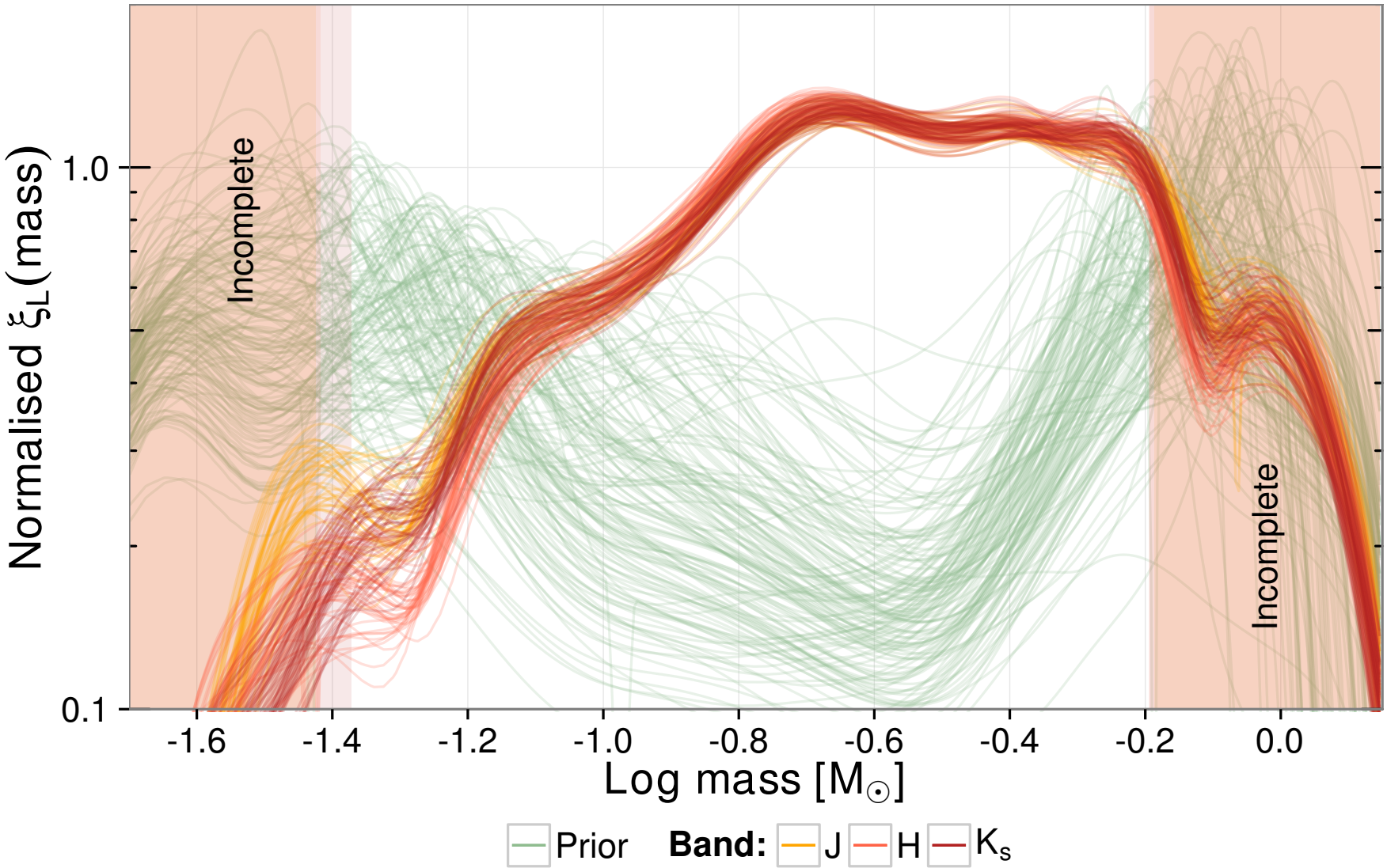
The statistical distributions: Empirical isochrone



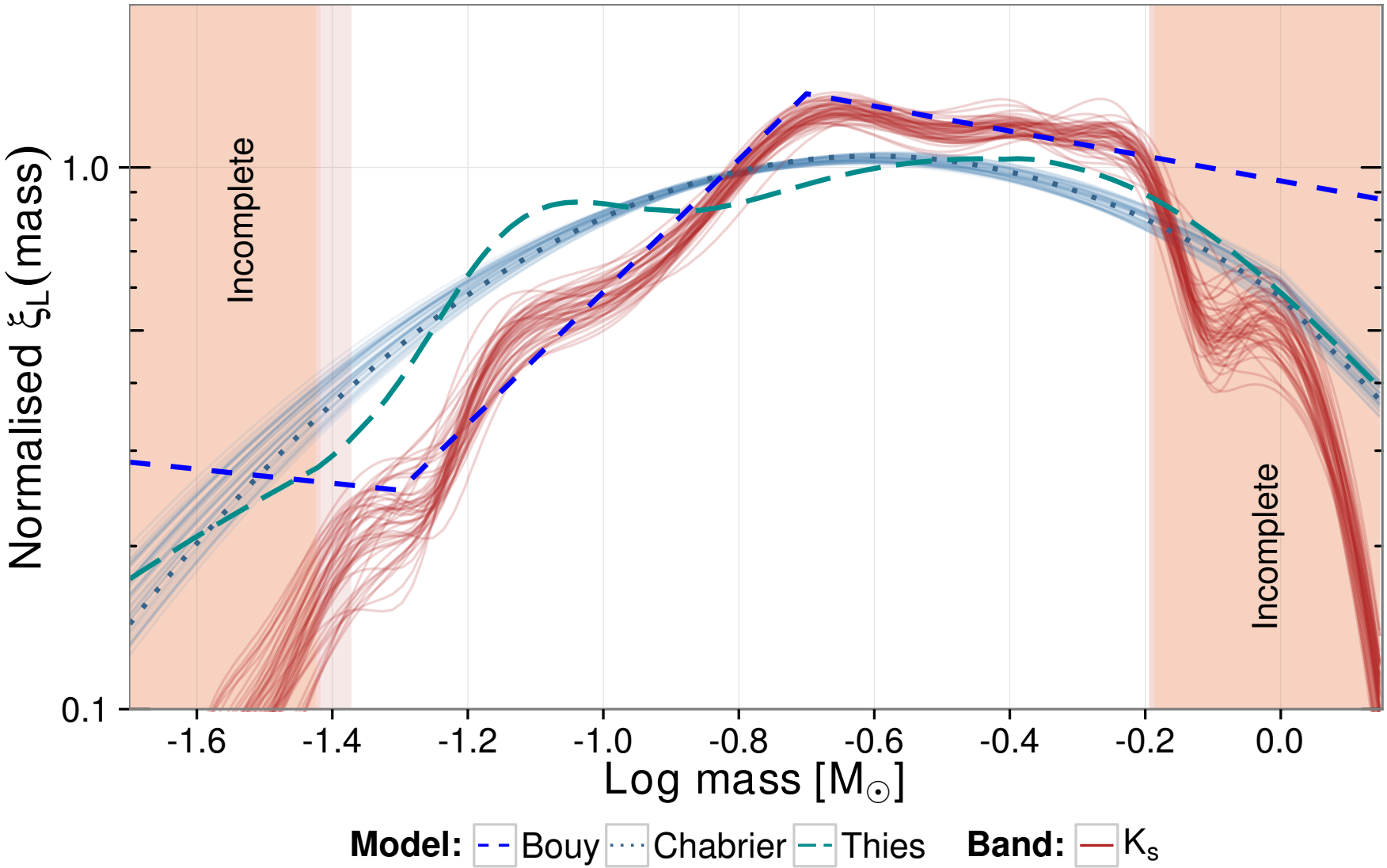
The statistical distributions: Luminosity



Results: The mass distribution



Results: The mass distribution



The Projected Spatial Distribution

Bayesian Model Selection

Models

- King 1962
- Elson et al. 1987
- Lauer et al. 1995
- Generalised King

Extensions:

- Radially symmetric
- Biaxially symmetric
- Luminosity segregated

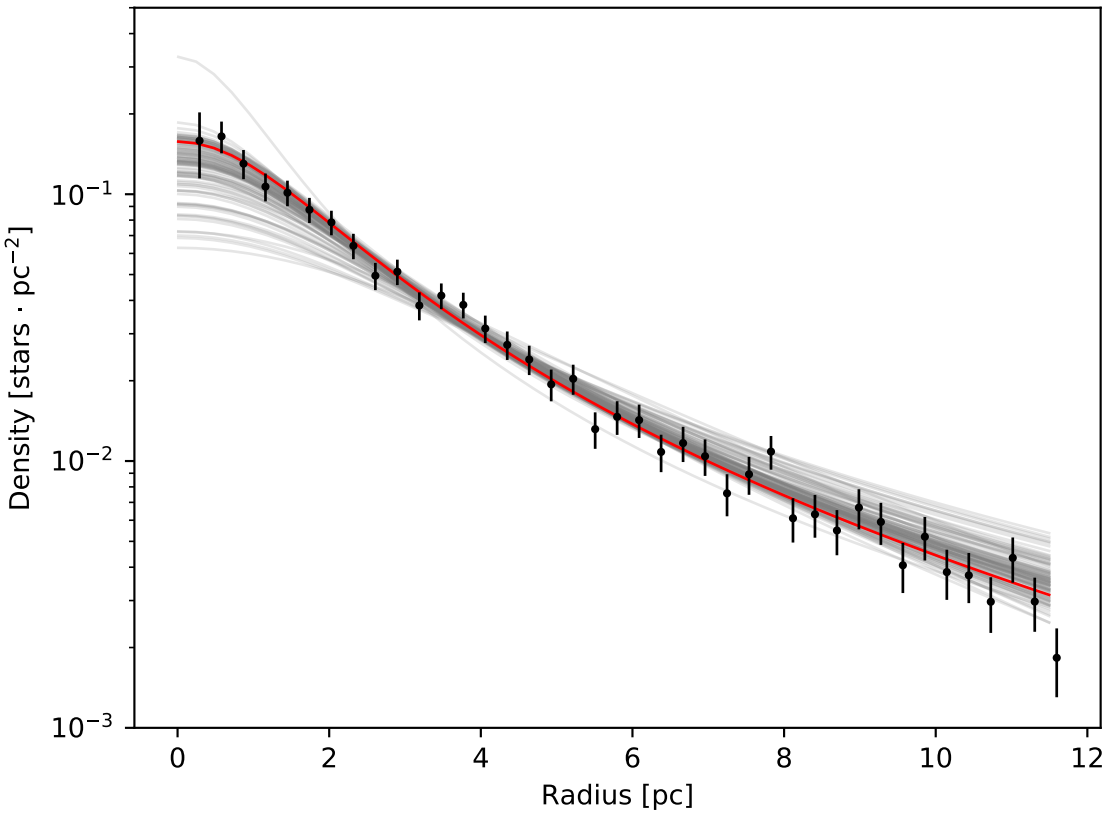
Data set

- The HMPS

Simultaneously

Bayesian evidence + Parametric inference

The Projected Spatial Distribution



Results

- Strong evidence:
  - Ellipticity
  - Luminosity segregation
- Best models
  1. King
  2. Generalised King
- Still lack 20% of members ($> 5°$)

Summary of results

- $80 - 90\%$ of candidate members in agreement with the literature
- 205 new cluster candidate members
- Mass distribution in discrepancy with IMFs in low-mass regime
- Projected Spatial Distribution
  - Evidence of ellipticity and mass segregation
  - One fifth of members beyond the survey area
  - Proved the validity of King's family of models

Conclusions

Objective fulfilled

- Statistical distributions of the cluster population
  - `Sky positions`
  - `Proper motions`
  - `Apparent magnitudes` $\rightarrow$ Luminosity $\rightarrow$ Mass
- Membership probabilities for 2×10^6 objects

- Constructed from 10^5 objects (missing values included)
- Non-homogeneous uncertainties propagated from all observables
- Full posterior distribution of the cluster

- Partial results submitted to A&A
  1. The Seven Sisters DANCe III: Bayesian Hierarchical Model
  2. The Seven Sisters DANCe IV: Projected Spatial Distribution

Limitations

- Field model
  - Does not include uncertainties
  - It is hold fixed $\rightarrow$ simultaneously inferred with the cluster
- Missing values
  - Not at random! $\rightarrow$ We need the detection probability
- Computing time $\rightarrow$ Parallelism in GPUs

Perspectives

- Apply to other clusters $\rightarrow$ COSMIC-DANCe $\rightarrow$ Luminosity distributions of Nearby Clusters
- Parallax, Radial velocities, Extinction, Binaries
- Data from: Pan-STARRS, Gaia, LSST, etc.
- Use empirical isochrones to improve stellar models.
- Liberate the software